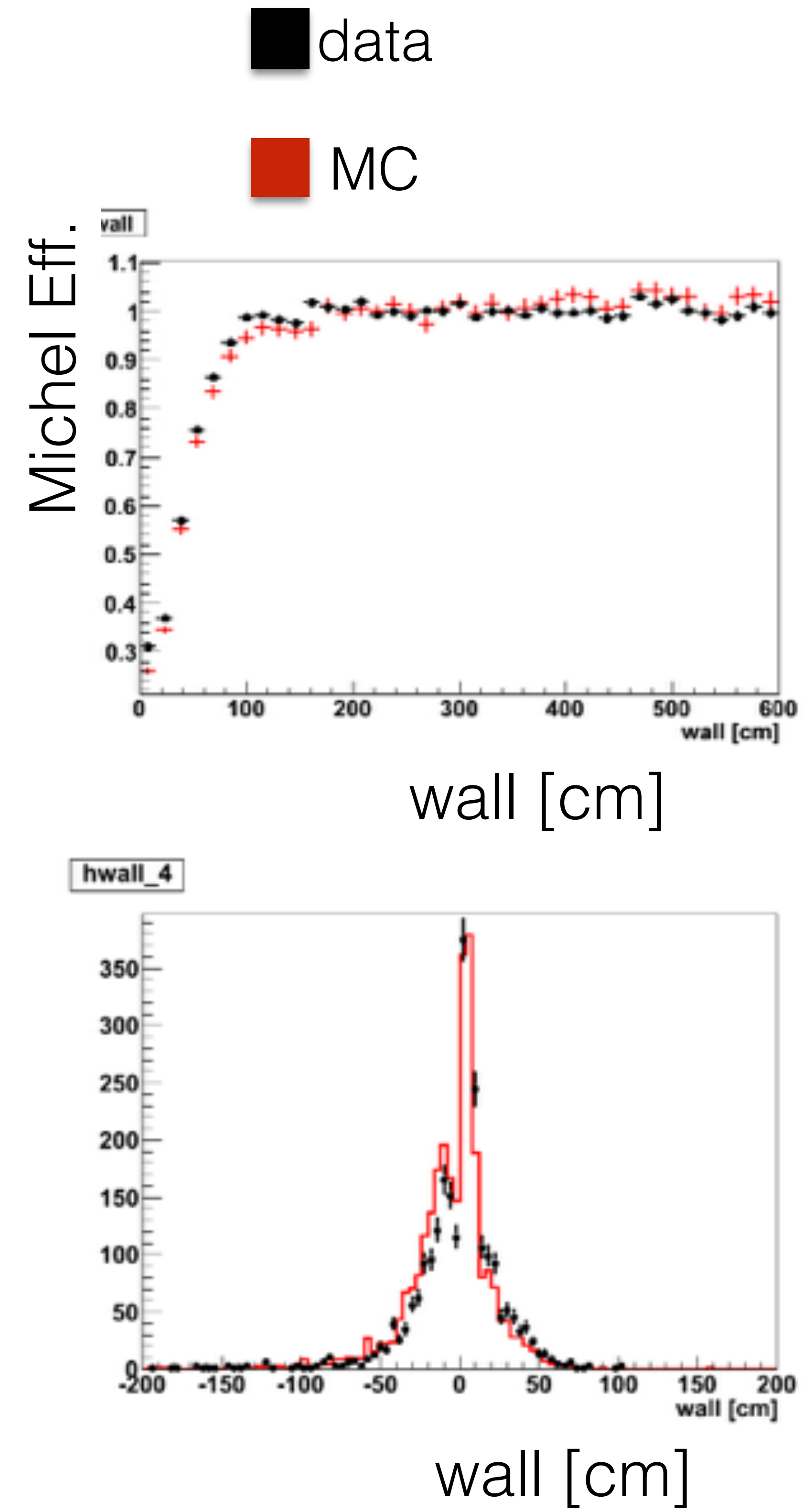
Approaches to Systematic Errors for T2K Fiducial Volume Studies

Andrew Missert
University of Colorado, Boulder
fiTQun! Workshop
August 2015

Introduction

- We need to use MC simulation to optimize fiTQun fiducial cuts
- Need to understand how well our simulation reproduces the data
- This means data/MC comparisons and propagation of errors to T2K neutrino energy spectrum
- Previous studies of data/MC differences (using stopping cosmic muons):
    - Estimate of data/MC difference in Michel tagging efficiency binned in **wall** variable
    - Estimate of data/MC difference in parameters governing the distribution of the **wall** variable for entering muons.
- These studies have focused on particular systematics that are especially well covered in the cosmic data



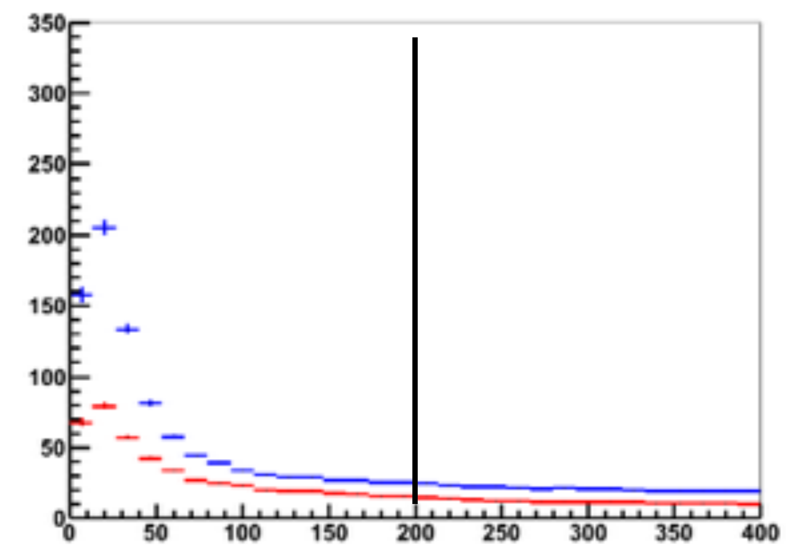
Introduction

- Clearly, there are many potential data/MC differences that cannot be checked using the cosmic muon sample
- The atmospheric neutrino sample covers many potential data/MC comparisons that are not possible with the cosmic muon sample
  - However, significant flux, cross section and true particle ID uncertainties make comparisons more difficult
  - Our estimation of detector systematic errors will have these uncertainties folded in
- We need a clear and comprehensive strategy for quantifying the data/MC differences in the atmospheric sample and propagating to uncertainties in the reconstructed T2K spectrum.
- This talk will outline some possible strategies:
  - 1) Extension of current atmospheric fit: fit MC expectations of events in various samples to data in each fiducial bin.
  - 2) Fit underlying distributions in MC to the corresponding data distributions
  - 3) Vary fundamental fiTQun inputs (scattering, absorption etc.) to cover data/MC differences

Potential Problems Near ID Wall

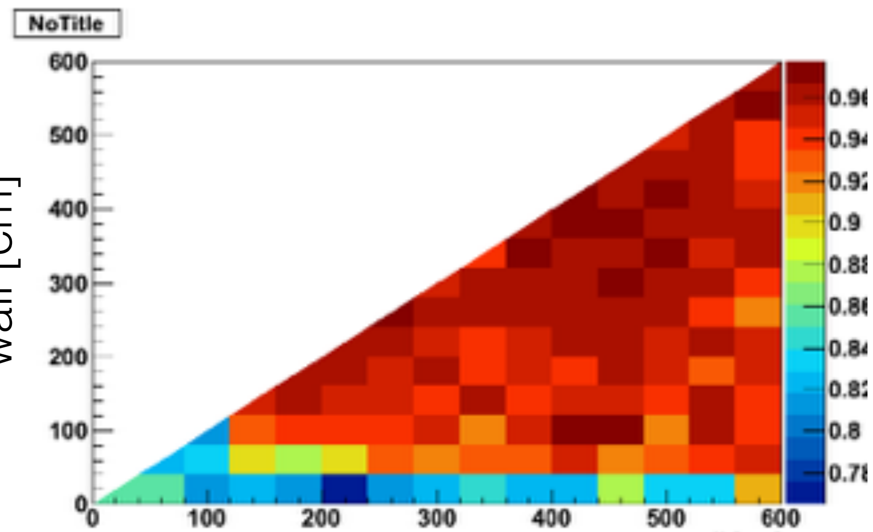
- Entering backgrounds from outside of ID
    - OD should reject most of these, but there are “dead region” events
    - Uncertainty in simulating how far into the detector we reconstruct outside events is well constrained by cosmic muon data.
- Degradation of reconstruction performance:
    - Momentum reconstruction: Few hit PMTs and/or deformed ring shapes resulting in larger momentum spread and bias toward high (or low) momentum.
    - Direction reconstruction: fiTQun has been shown to be unreliable in reconstructing direction at vary small **wall** and (true) **towall** values. Track has larger direction spread and is sometimes wildly mis-reconstructed in nearly the opposite direction.
    - PID discrimination: Small rings and few hit PMTs decrease the likelihood differences that we use for PID
    - Ring counting: Tend to lose rings from MR events **and** gain rings in single ring events.
- **Our systematics analysis strategy should be sensitive to how well we simulate these issues**

average Δ p [MeV]



wall [cm]

Ring Counting Cut Eff.



wall [cm]

towall [cm]

Extending Current Analysis

- One method of estimating systematic error is to extend current systematic error analysis to other fiducial cut bins

- Current strategy at SK: maximize L(**x**|**b**,**σ**,**ϵ**) where

  - **x** = N_{ijk} , the number of events in reconstructed sample k in kinematic bin i,j

  - **b** = flux uncertainty parameters

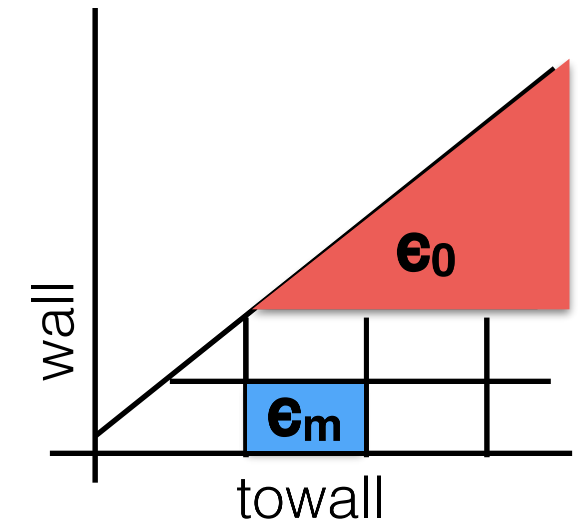
  - **σ** = cross section uncertainty parameters

  - **ϵ** = parameter that adjusts MC expectation for each true interaction mode in each bin

- Could either extend current framework or build a new one that operates on the same principal

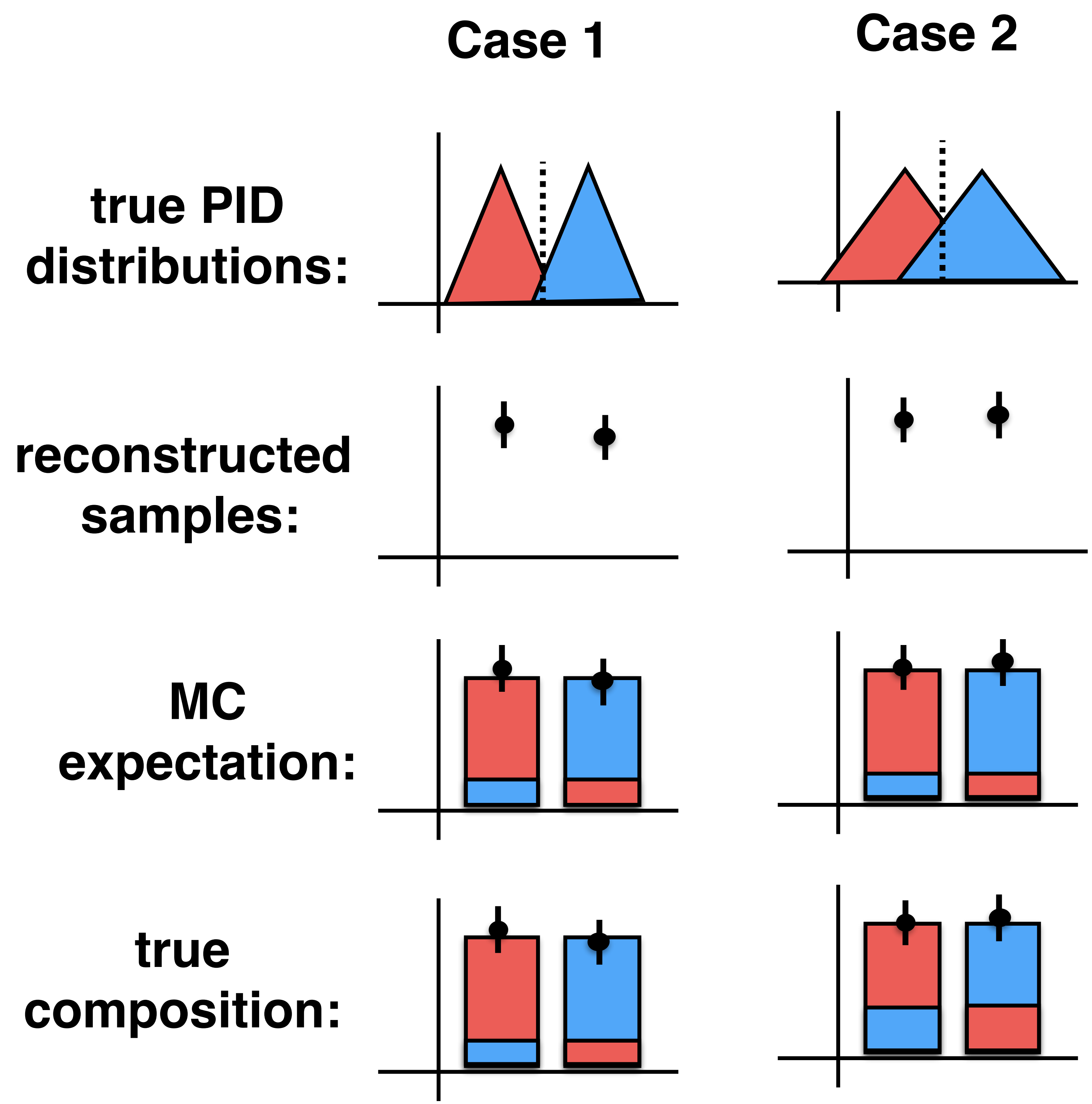
Extending Current Analysis

- To extend current analysis just repeat the likelihood maximization in each fiducial cut bin
  - If we want to add m fiducial bins, then add m copies of the $\boldsymbol{\epsilon}$ parameters and maximize $L(\mathbf{x}_m|\mathbf{b},\boldsymbol{\sigma},\boldsymbol{\epsilon}_\mathrm{m})$ where $\mathbf{x}_m = N_{ijkm}$, the number of events of sample k in kinematic bin i,j in fiducial region m.
- Statistical errors may be problematic. (~100 events in each fiducial bin)
  - Would probably need coarser kinematic binning and/or broader event samples
- $\boldsymbol{\epsilon}$ is already a large vector of parameters, we would be increasing the number of parameters in the fit by the number of fiducial bins (~10)



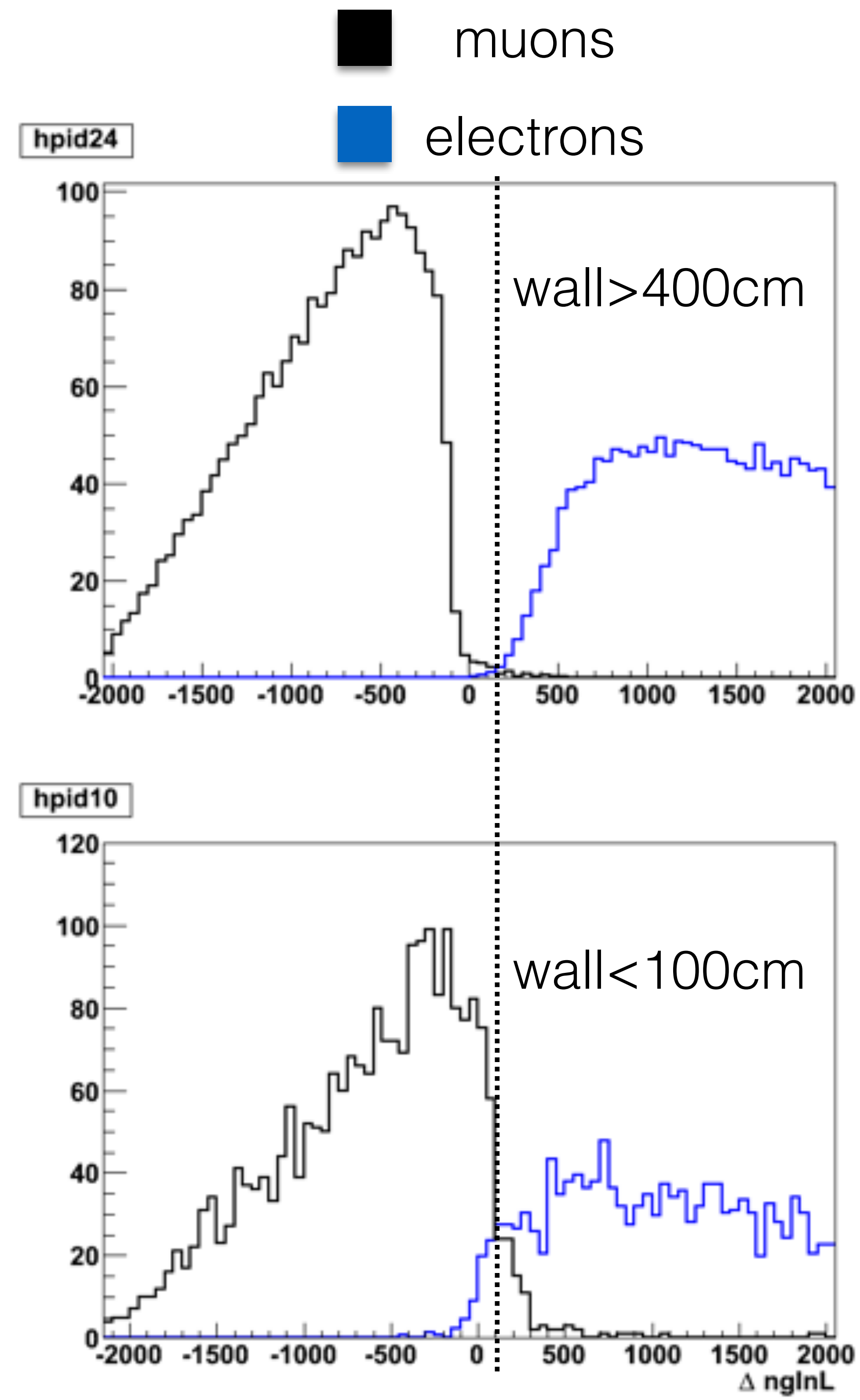
Extending Current Analysis

- This type of analysis may be blind to differences in resolution between data and MC
    - Cannot discriminate between case 1 and case 2 in this example
- This is of particular concern since this is exactly the type of situation we want to be sensitive to



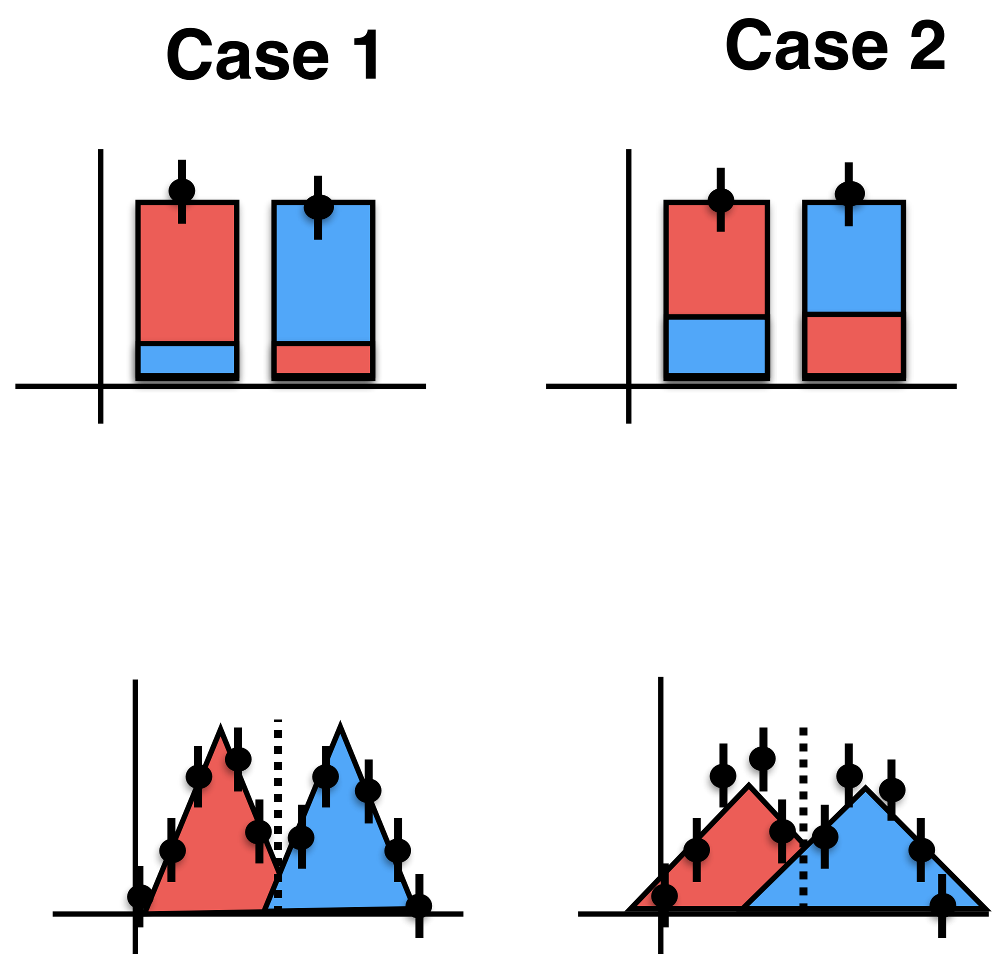
Extending Current Analysis

- One of the important issues near the ID wall is the decreased resolution of reconstructed quantities
  - For example, the overlap of likelihood distributions for electrons and muons increases as we get closer to the wall
- Extending the current approach to systematics may not tell us much about how well we simulate this smearing of likelihood distributions
- Alternative strategy, compare underlying distributions between data and MC, propagate errors to N_{SK}



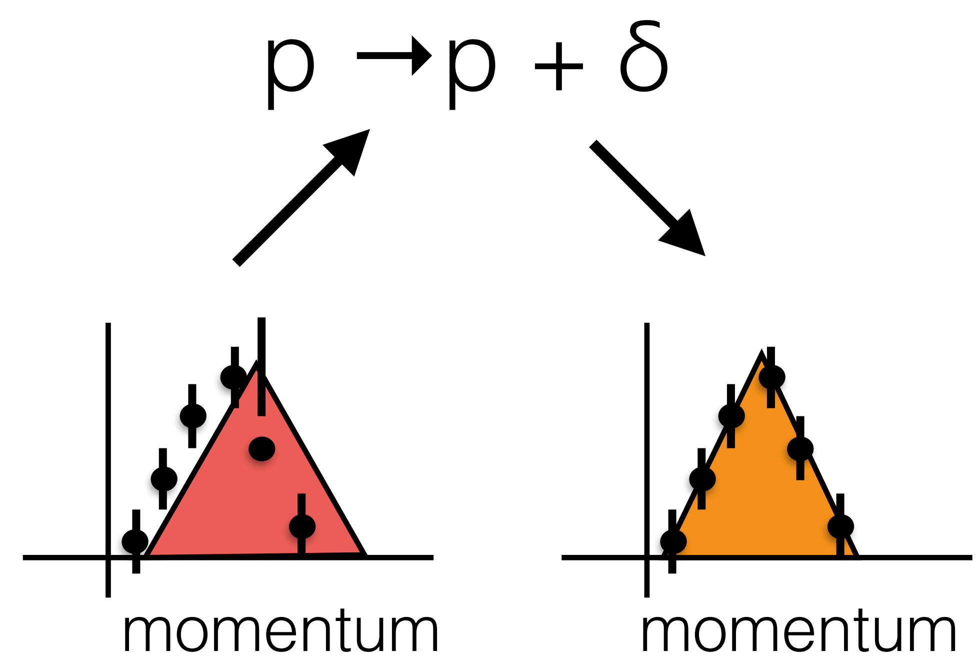
Fitting to Reconstructed Distributions

- How can we differentiate between these two cases?
  - Check shape of underlying distributions directly
- Data/MC differences in resolution (and bias) can then be propagated using toy MC
- Need a way to parameterize MC expectation and fit to data
  - Directly fit for resolution smearing and bias



Fitting to Reconstructed Distributions

- We want to check for differences in bias and resolution between data and MC, so parameterize MC expectation using these quantities
  - For example, assume we would like to check the MC prediction assuming a bias β and a resolution ρ
  - For each event in the MC we change the reconstructed momentum: $p \rightarrow p + \delta$, where δ is drawn from an normal distribution of mean β and width ρ.
  - Re-fill momentum histogram and compare to data using some metric (chi-squared)
  - Maximize metric over (β,ρ) to get best momentum and bias assumptions
- Practically, this way of fitting (β,ρ) for each distribution would be too slow, would have to refill histograms for each guess of (β,ρ)



Fitting to Reconstructed Distributions

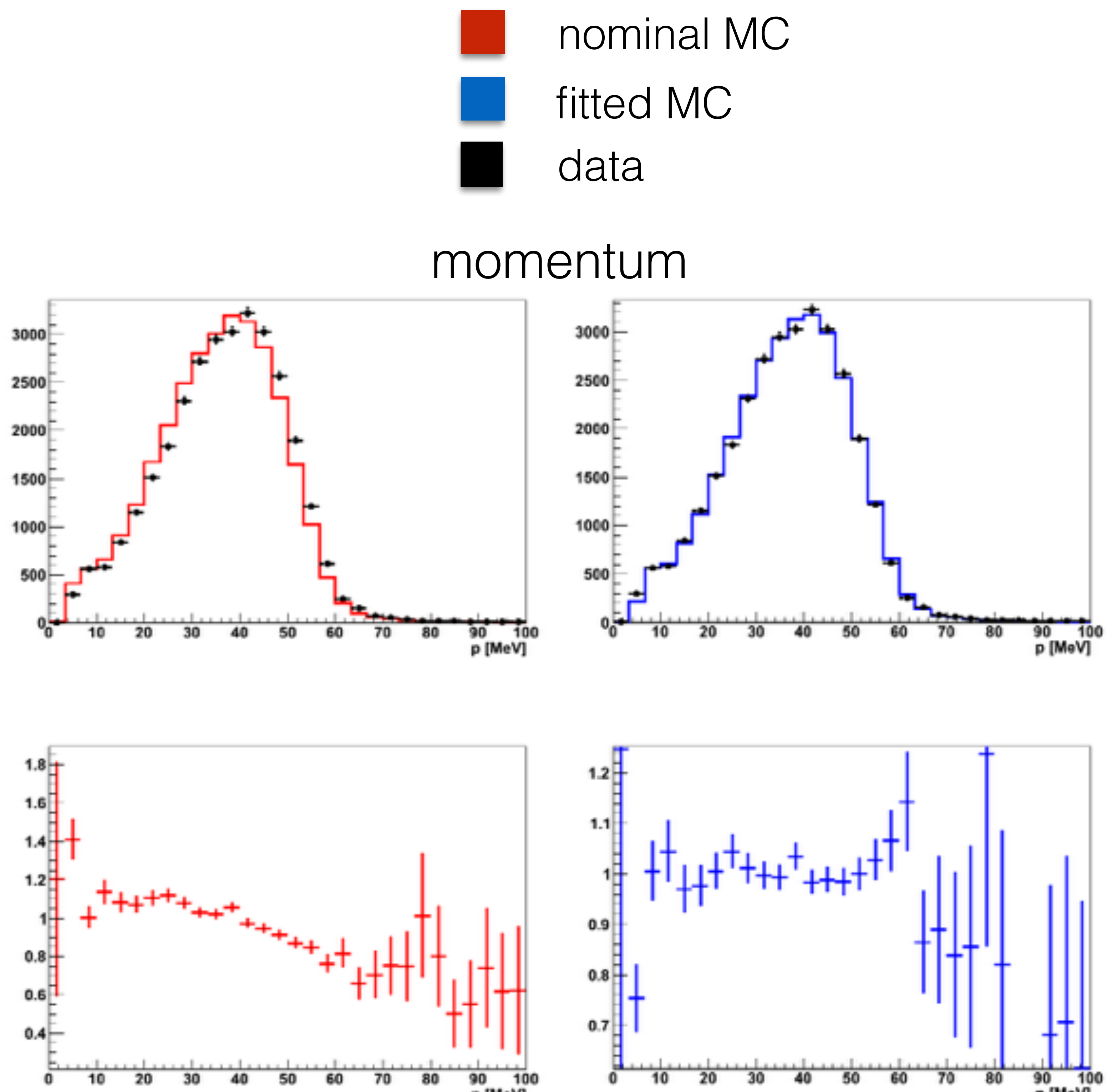
- Need a way of quickly calculating the expected distribution for an assumption of (β,ρ) given the nominal MC distribution
- **Proposition**: Let X be an random variable with PDF f(x). Let △ be a random variable with Gaussian PDF *G(δ;μ,σ)*, the the PDF *g(y)* for the random variable *Y = X +* △ is given by:

  g(y) = (f*G)(y)

  - where "*" denotes convolution of the two functions.
- For the problem at hand, this means that we can get a new MC prediction by just convolving the nominal histogram with a Gaussian G(x;β,ρ)
- Can now maximize likelihood using MINUIT. Very fast.

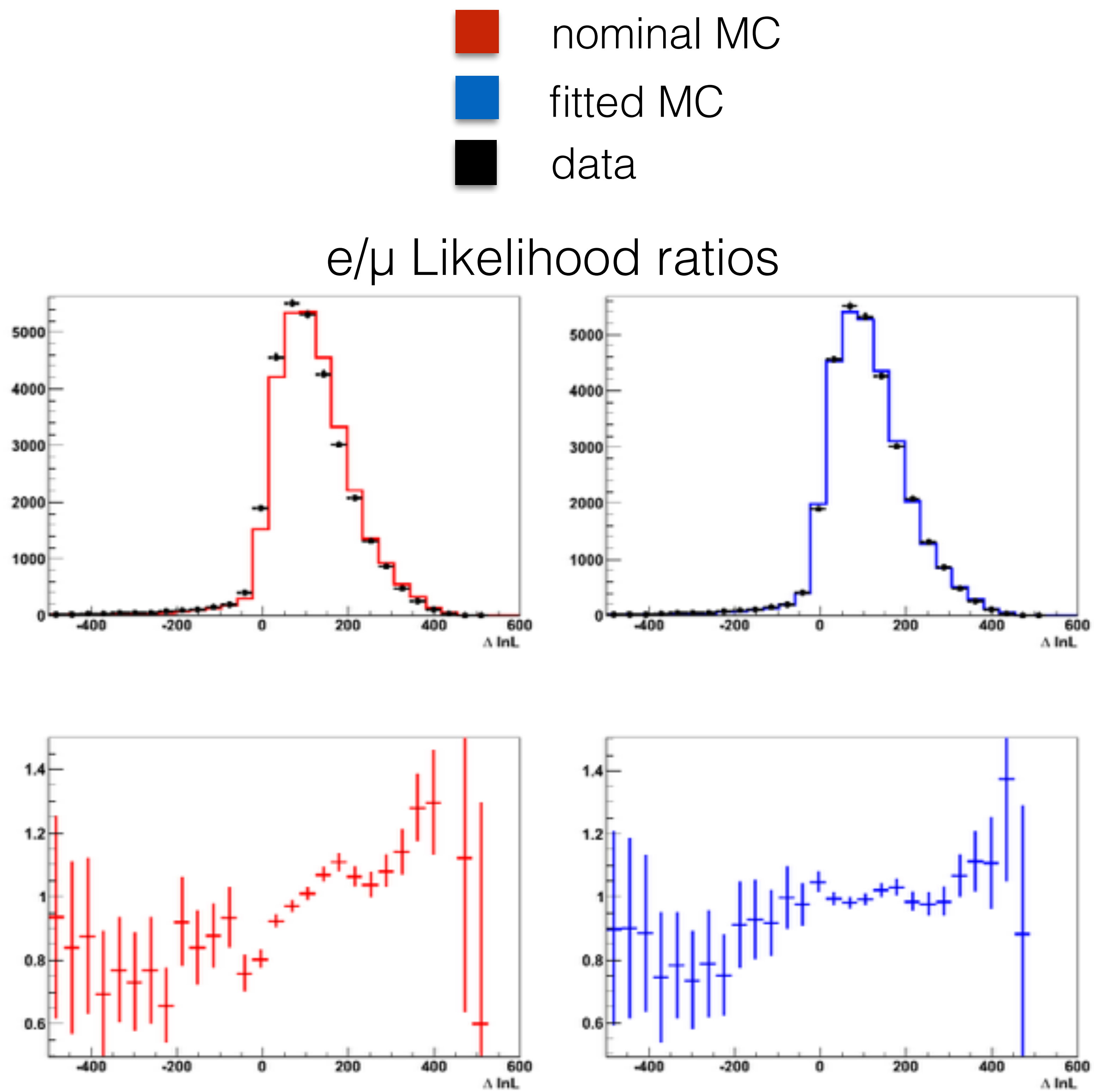
Fitting to Reconstructed Distributions

- Example: fit to data/MC difference in Michel electron spectrum for stopping cosmic muons:
  - Cuts:
    - fqtotq>100
    - fqnse=2
    - fqpcflg[1][1]=0
- Best fit: resolution = 1.26 MeV, bias = 1.17 MeV
- Difference reduced from 15-20% to 5-10% in most bins



Fitting to Reconstructed Distributions

- Example: fit to data/MC difference in Michel electron spectrum for stopping cosmic muons:
  - Cuts:
    - fqtotq>100
    - fqnse=2
    - fqpcflg[1][1]=0
- Best fit: resolution = 4.9, bias = -6.9
- Difference reduced from 15-20% to 5-10% in most bins



Fitting to Reconstructed Distributions

- To apply this to atmospheric error analysis use approach similar to what was described previously:
  - Separate data into *m* fiducial cut bins
  - In each bin, fill histograms of fiTQun reconstruction outputs for various samples (PID, momentum, nring etc.)
  - Fit assumptions of bias and resolution ($\boldsymbol{\beta},\boldsymbol{\rho}$) in each fiducial cut region
    - Maximize $L(\mathbf{x}_m|\mathbf{b},\boldsymbol{\sigma},\boldsymbol{\beta}_m,\boldsymbol{\rho}_m)$ across all *m*
  - Propagate errors to T2K neutrino spectrum using toy MC

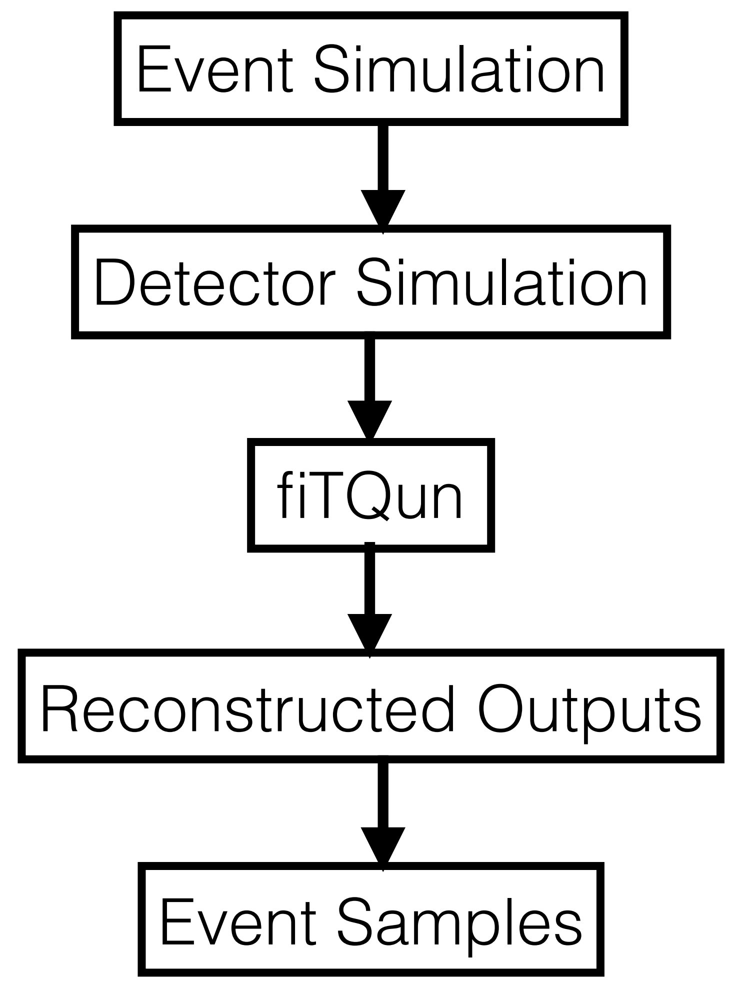
Fundamental Approach

- Our (detector) systematic errors essentially come from how well we model out detector
- Potential source of error in simulation:
    - Water quality
    - PMT response
    - Geometry
- Ideally we should be able to produce uncertainties in higher-level quantities by varying these fundamental parameters
    - Only use atmospheric analysis to check that we can cover all data/MC differences
    - No folding atmospheric flux uncertainties into detector uncertainties
- Unfortunately, SKDETSIM is not set up to easily do this
    - Long-term project
- Can we “cheat” and vary fiTQun parameters instead?
    - Change weights of scattered vs. direct charge
    - Modify scattering tables, assumptions of PMT response, water transmission factor

Summary

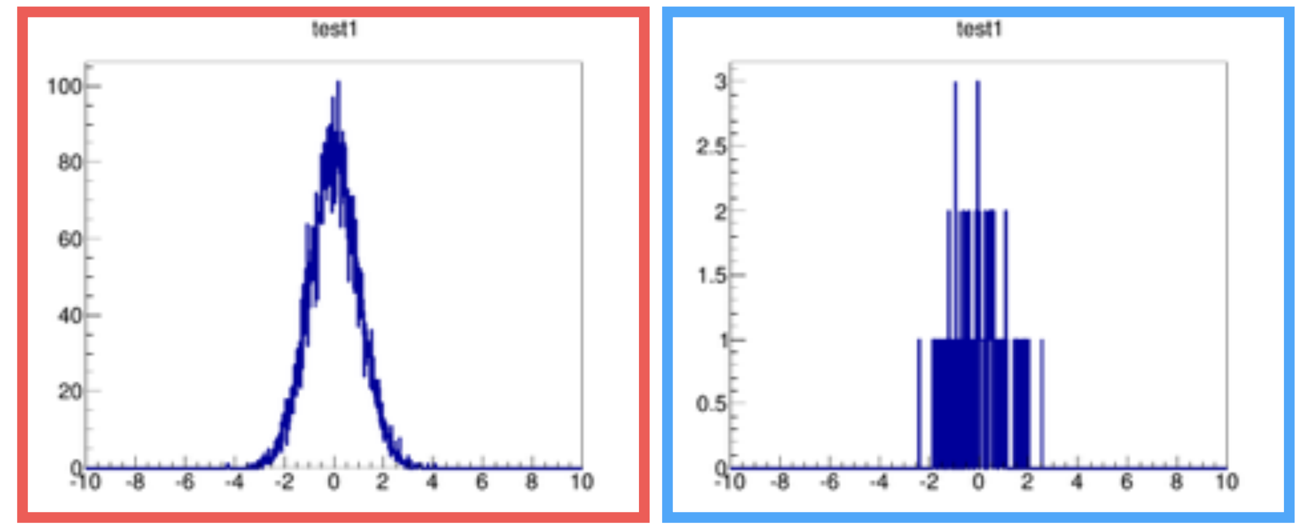
- Choice of strategy essentially comes down to where in the simulation chain we modify/fit the MC expectations
- Current analysis fits MC expectations for various event samples
  - May not be sensitive to differences in resolution between data and MC
- Fitting our assumptions of bias and resolution to the data seems like a more direct approach
  - Gaussian smearing method allows for quick fitting of bias and resolution parameters between data and MC
  - Event-by-event variable modification provides way of propagating errors forward in the analysis
  - Short-term approach
- Variation of parameters at detector simulation is most appealing approach, but is long term solution
- Variation of fiTQun water quality assumptions may be short or long term

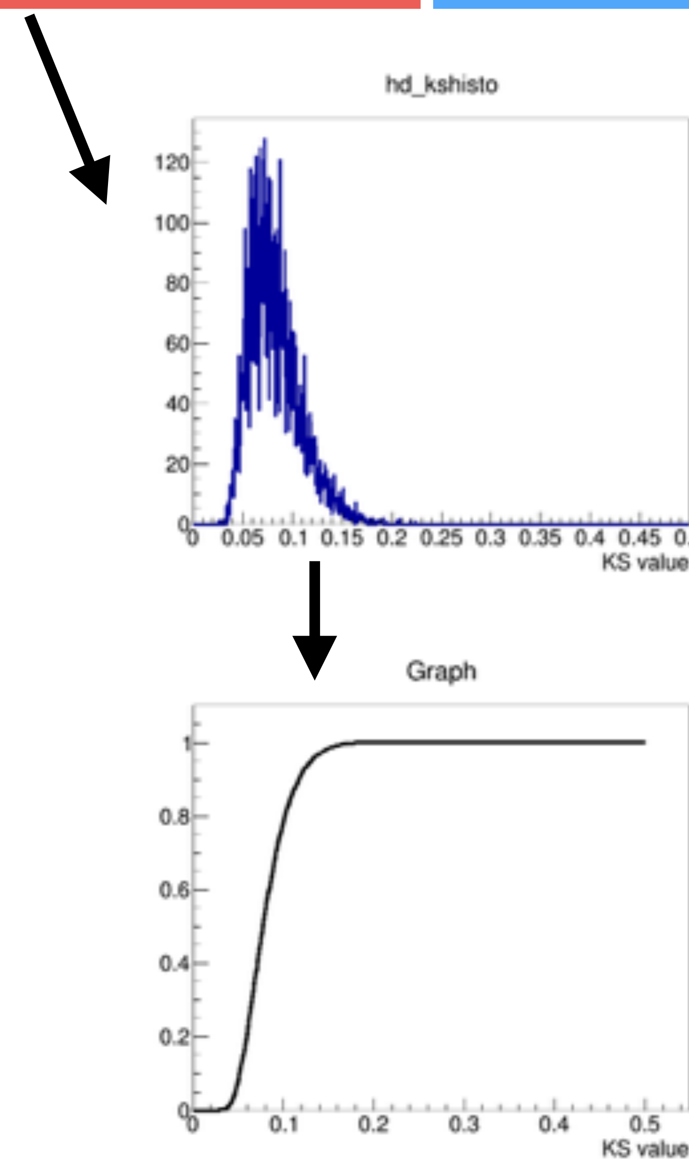
Where do we modify MC expectation?



KS Probability

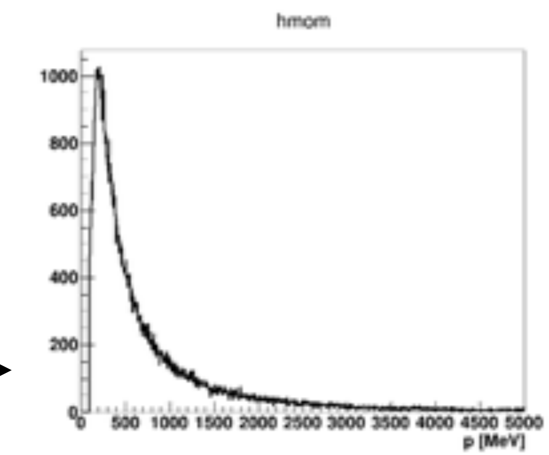
- How to calculate the Kolmogorov–Smirnov probability:
  - Compare histogram h1 with entries N1 with histogram h2 with entries N2
  - The KS test statistic “k” is given by the maximum difference between the CDFs of h1 and h2
  - How do we assign a probability to a particular value of k?
    - Use MC technique: throw M (~10000) random histograms **of size N2** from histogram h1
    - Make distribution of k for these pseudo-experiments
    - 1 - CDF(k) gives a p-value for k (0.65 in this case)



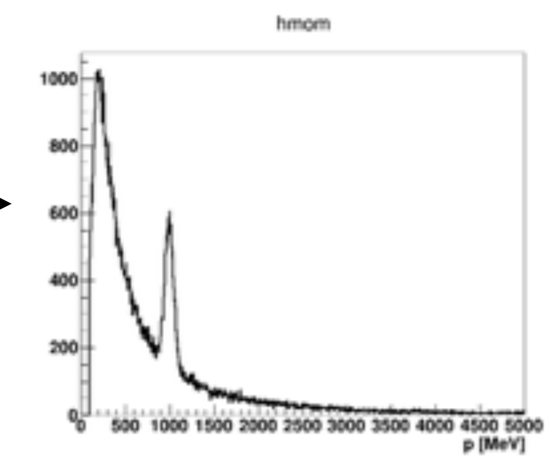


KS Sensitivity

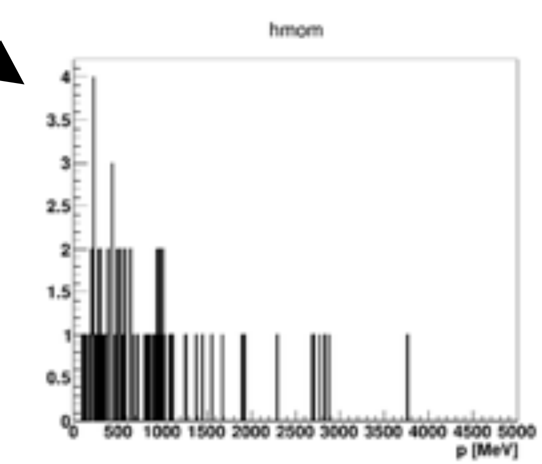
- KS sensitivity test:
  - Take histogram of reconstructed momentum for atmospheric e-like MC events with wall>200. This is “core” sample histogram.



  - Perturb this histogram with a “bump” of width 50 MeV (slightly larger than momentum resolution). The size of the bump is given by a number “x” s.t. if N is the number of events in the histogram, x*N is the number of events in the bump



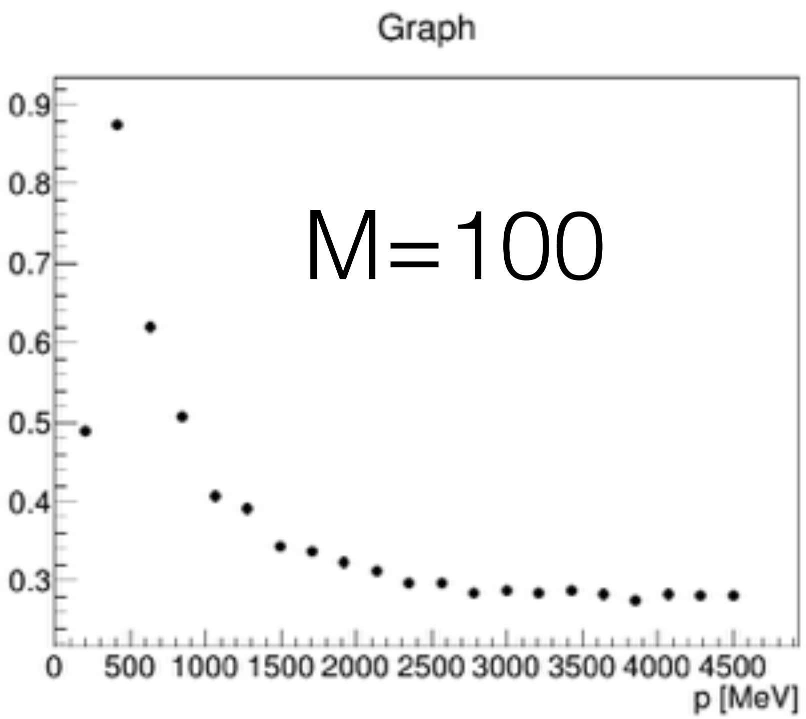
  - Throw “M” entries from the perturbed histogram and evaluate the KS probability. Repeat this ”L” times to get a set of L KS probabilities.

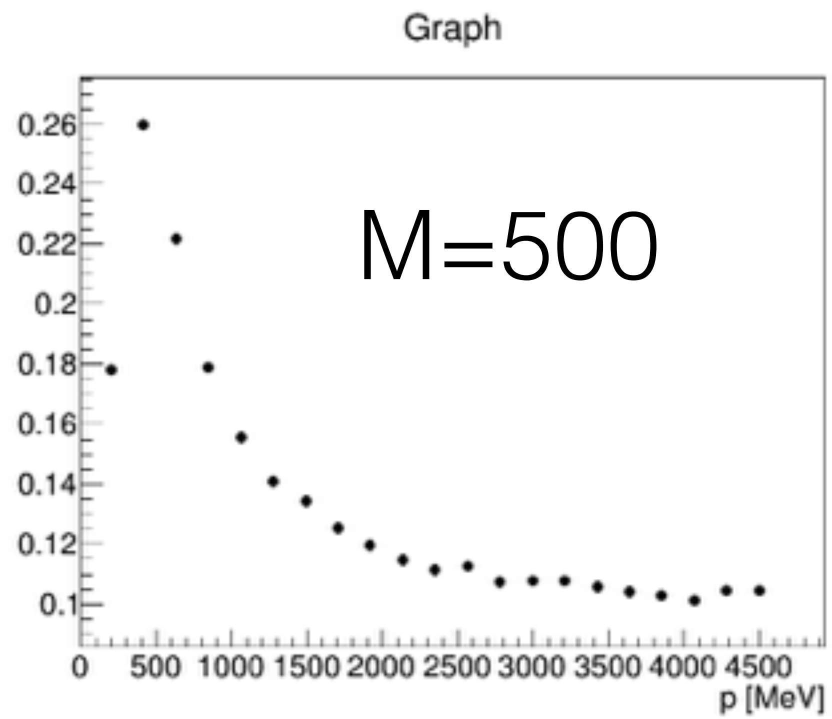


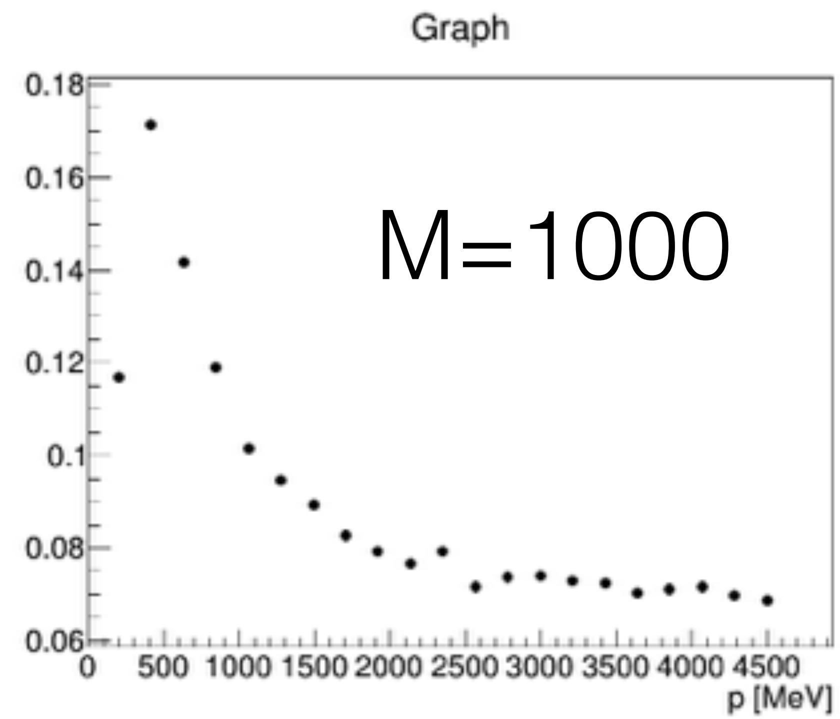
X **L**

  - If less then 90% of these L experiments pass the KS test at significance α, try a larger bump
  - Else: record the size of the bump and repeat the process for a bump at a different momentum.

KS Sensitivity



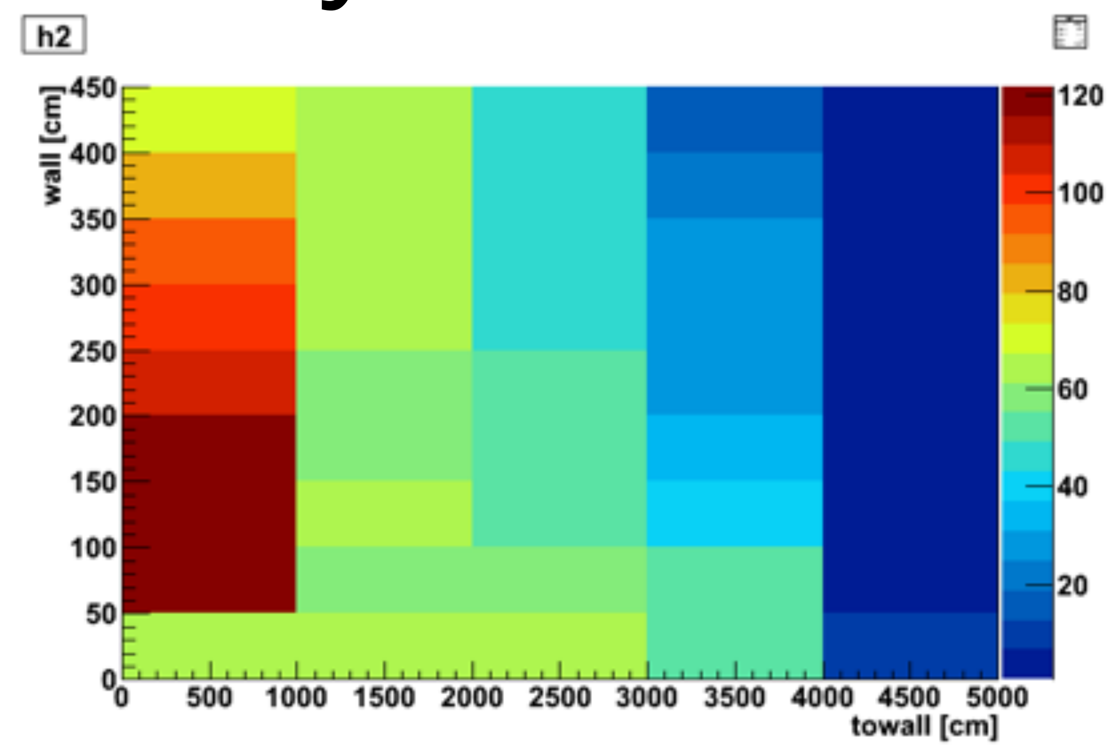


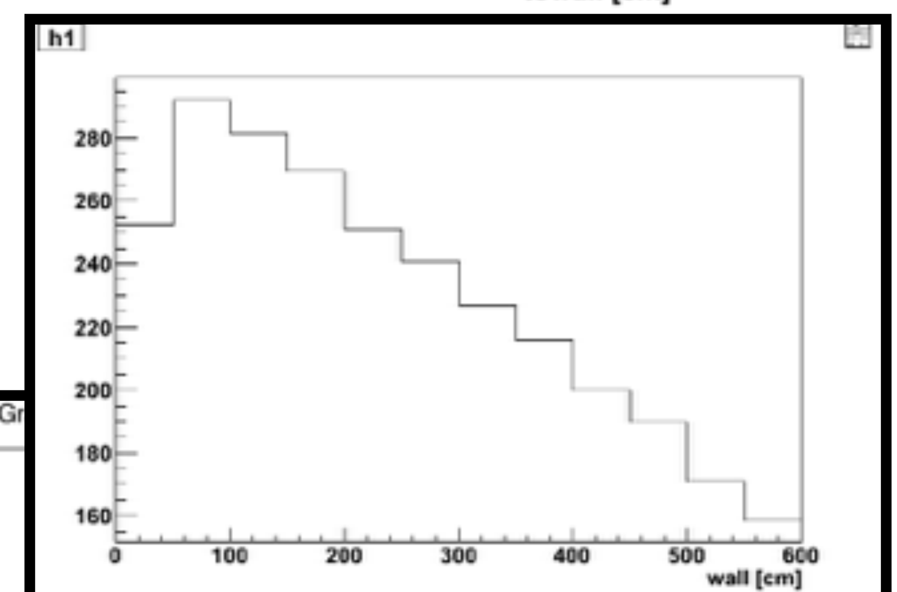


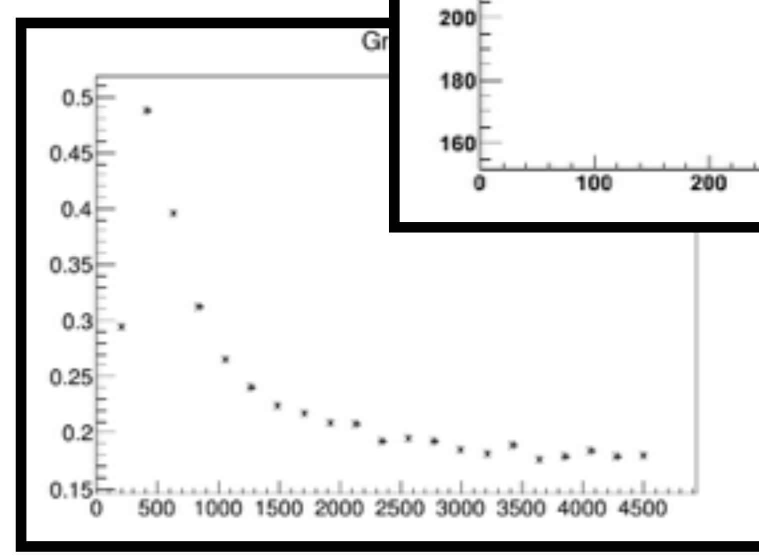
- Unfortunately, the KS test is not that sensitive for low statistics comparisons
- For this momentum spectrum, you are only sensitive to bumps 90% the size of the original histogram in some momentum regions if you are comparing to a histogram with 100 events
- The sensitivity increases sharply as the number of events in the histogram you are comparing increases
- Need >1000 events to be sensitive to shape difference in the 10% range.

KS Sensitivity

- Do we have enough events in the atmospheric data to make a significant statement about the fiducial volume?
- For example, there are only ~4000 e-like sub GeV events in the atmospheric sample
- For corse binning in fiducial cut variables {wall, towall} ~100 events per bin
- For 1D cut, ~200 events per bin



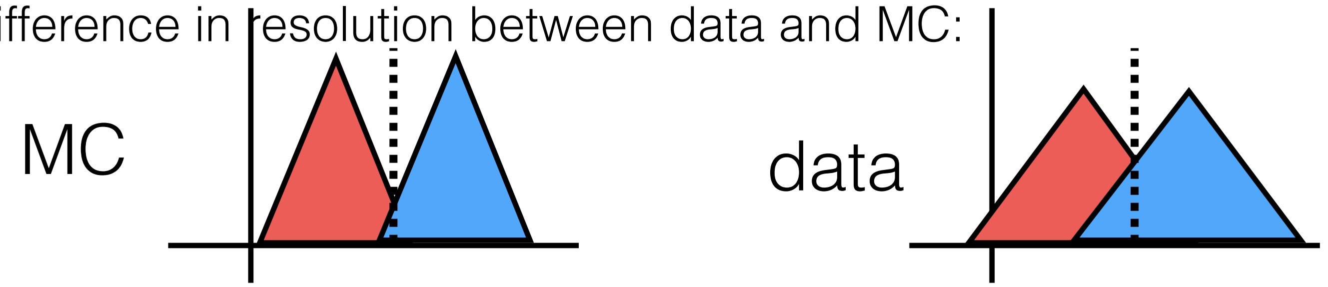




	SK-I		SK-II		SK-III		SK-IV	
	Data	MC	Data	MC	Data	MC	Data	MC
FC sub-GeV single-ring								
e-like								
0-decay	2992	2705.4	1573	1445.4	1092	945.3	2098	1934.9
1-decay	301	248.1	172	138.9	118	85.3	243	198.4
π^0-like	176	160.0	111	96.3	58	53.8	116	96.2
μ-like								
0-decay	1025	893.7	561	501.9	336	311.8	405	366.3
1-decay	2012	1883.0	1037	1006.7	742	664.1	1833	1654.1
2-decay	147	130.4	86	71.3	61	46.6	174	132.2
2-ring π^0-like	524	492.8	266	259.8	182	172.2	380	355.9

Other Approaches

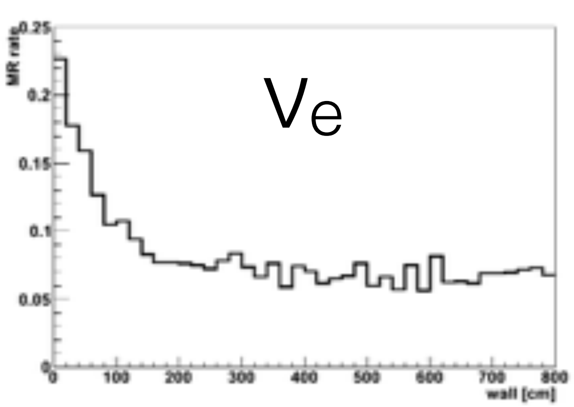
- If we use MC to optimize fiducial cuts, we need to estimate systematics using the atmospheric sample
- What about a fit the atmospheric data similar to what is currently done for T2K?
- Essential principal of atmospheric fit: minimize N_{MC} - N_{data} where N_{MC}(flux and xsec uncertainties, efficiency corrections)
  - Result is constraint on efficiency corrections
  - Differences in migrations between samples between data and MC should be well constrained here
- Can we use this technique for FV systematics?
  - Perform fit in each bin in {wall,towall} space
- Potential issue: difference in resolution between data and MC:

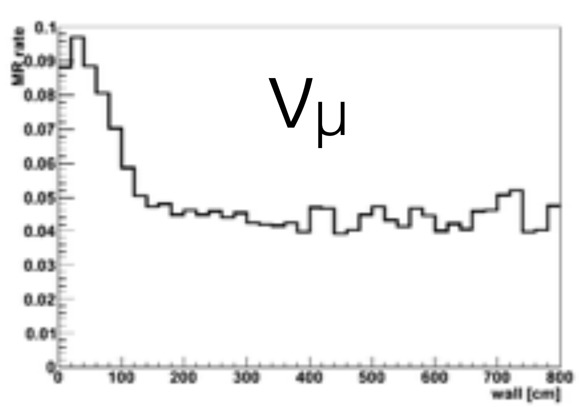


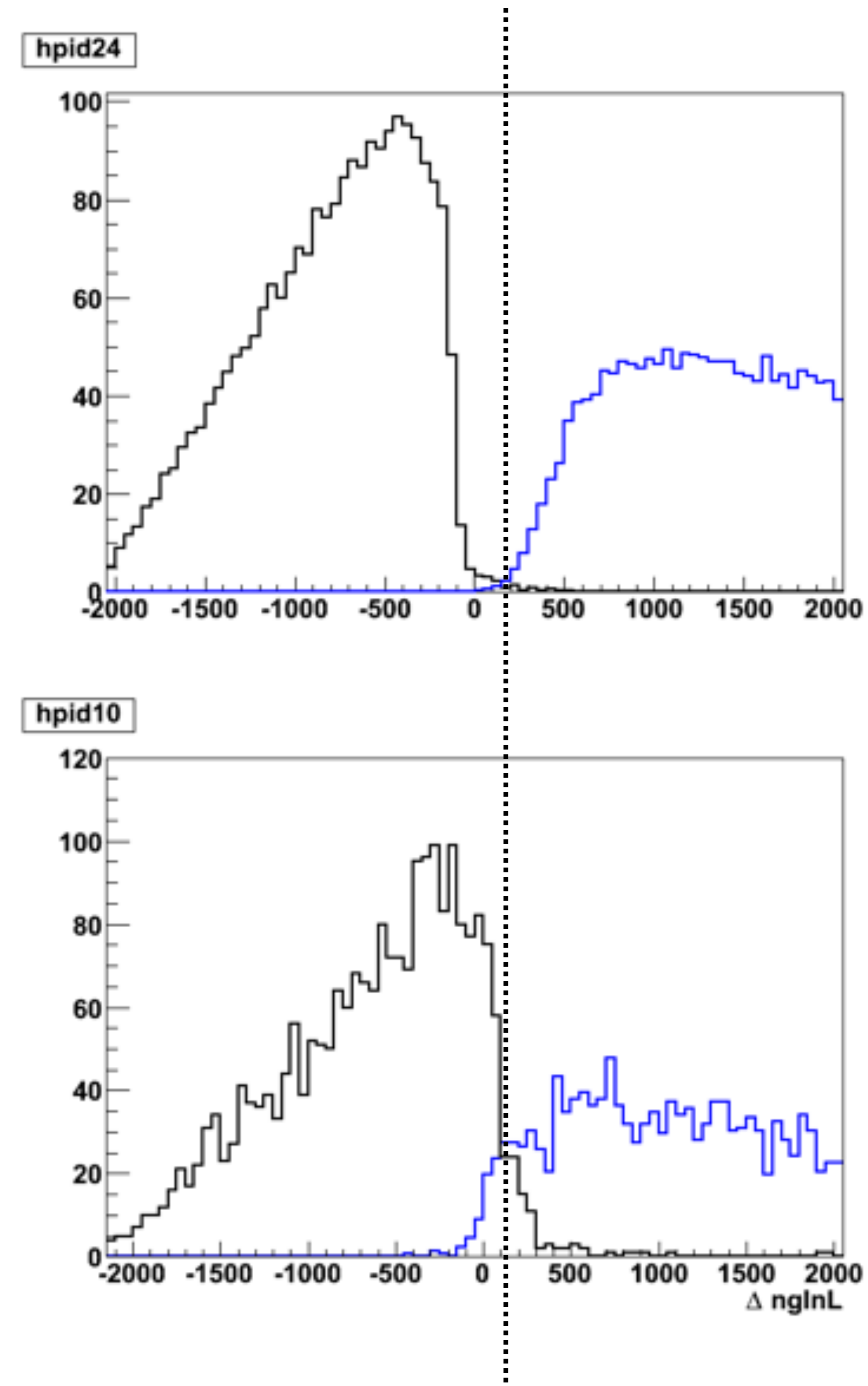
  - Numbers of events for each sample are similar for data and MC → small shift error → underestimate # of BG events in fiducial cut region

Other Approaches

- Differences in resolution are of particular importance for fiducial volume systematics.
- One example: PID near ID wall
- Shapes of likelihood distributions spread out for both muon and electron events near the ID wall
    - Optimal cut value barely changes
- Data/MC differences in this decrease of PID resolution may be masked by just fitting N_{SK}
- Similar effect for the resolution of the number of rings:



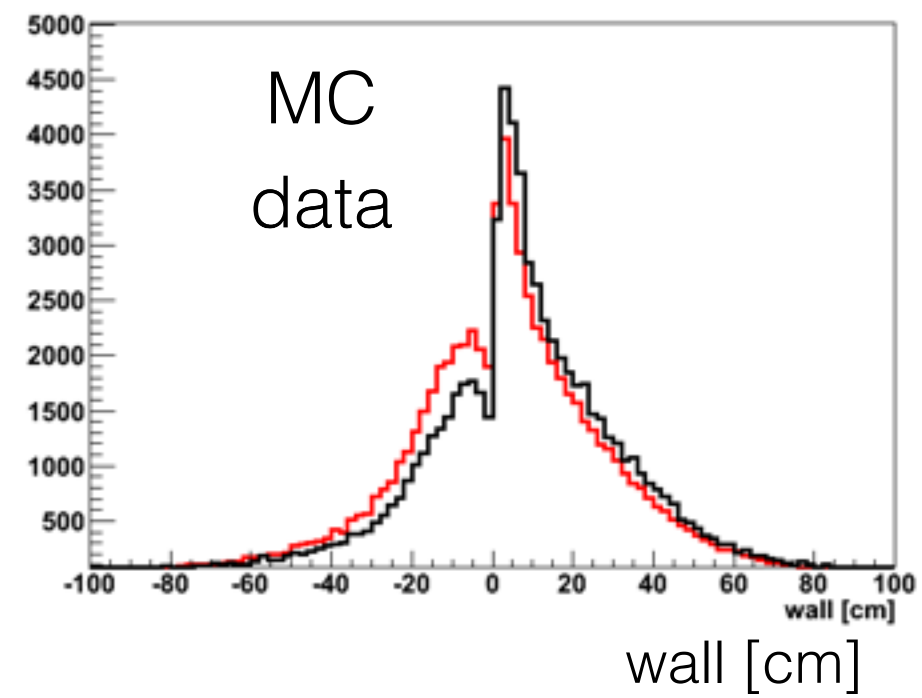




Other Approaches

- Estimating systematic uncertainties in resolution can be done by checking data/MC differences in the underlying distributions
  - Similar to what I've done with the differences in the wall distribution for cosmic muons
  - Uncertainties in how the distribution is modeled are then propagated forward using reweighing or observable variation
  - Potential issues: Have to find parameterization for data/MC differences
- Another possibility: Fit for relative efficiency
  - Instead of fitting differences in N_{SK} between data and MC, fit differences in NSK between pairs of data samples: a "core" sample from the center of the tank, and a "test" sample from a fiducial cut region.
  - Once again, statistics may limit how well we can fit the relative efficiency

Stopping cosmic muons:



Conclusions

- If we want to be sensitive to ~10% changes in the spectrum shape, we don't have enough events in the atmospheric sample to determine a consistent fiducial volume with a resolution <50 cm
- Some possible ways to improve this analysis:
    - ~~Anderson-Darling Test: Uses more information than the KS test and is more sensitive~~ (only slight increase in sensitivity in tail of distribution)
    - Include normalization: Will be more sensitive to migrations between subsamples.. Have to be careful to normalize the effective fiducial volume element properly
    - It's likely we will have to rely on MC simulation to optimize the FV cuts. Use fit to atmospheric data to determine systematics.
- Approaches to FV systematic errors using atmospherics:
    - NSK fit similar to what is already being done for T2K
    - Data/MC difference parameterization for underlying cut distributions, propagate uncertainties to NSK
    - Relative efficiency fit between atmospheric data samples in various fiducial regions
- We should select an approach following some more quantitative studies and discussion at the fiTQun workshop